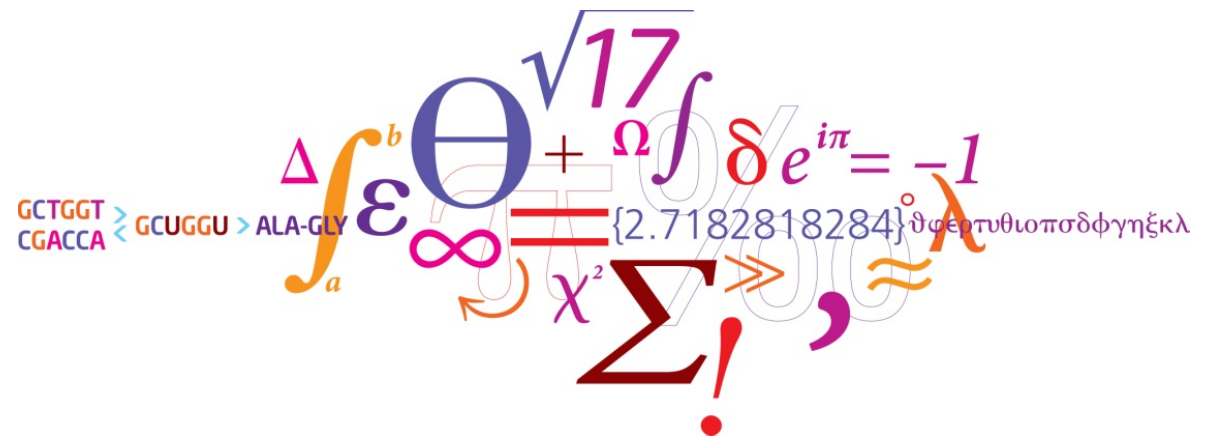
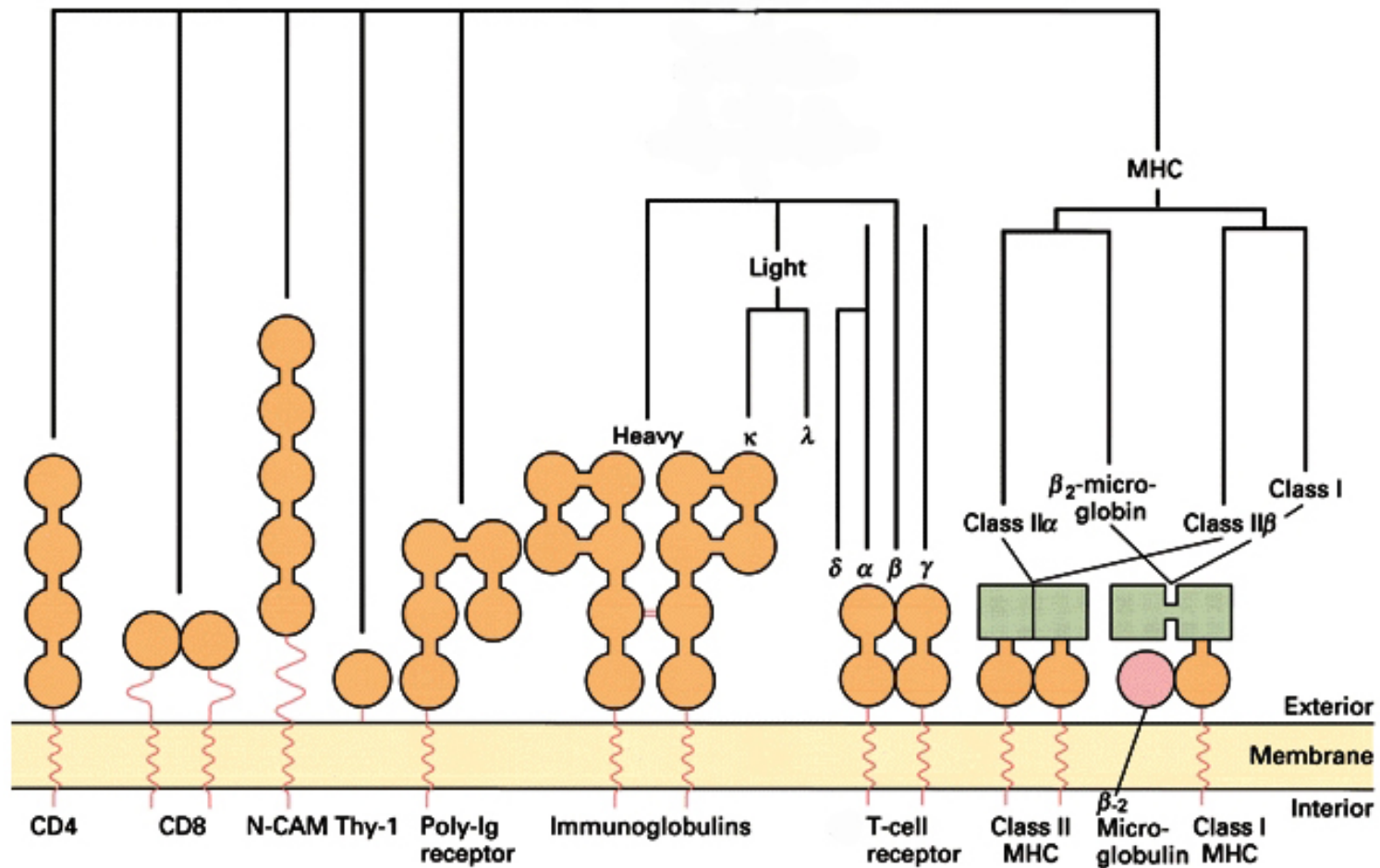


# Antibody and TCR structure

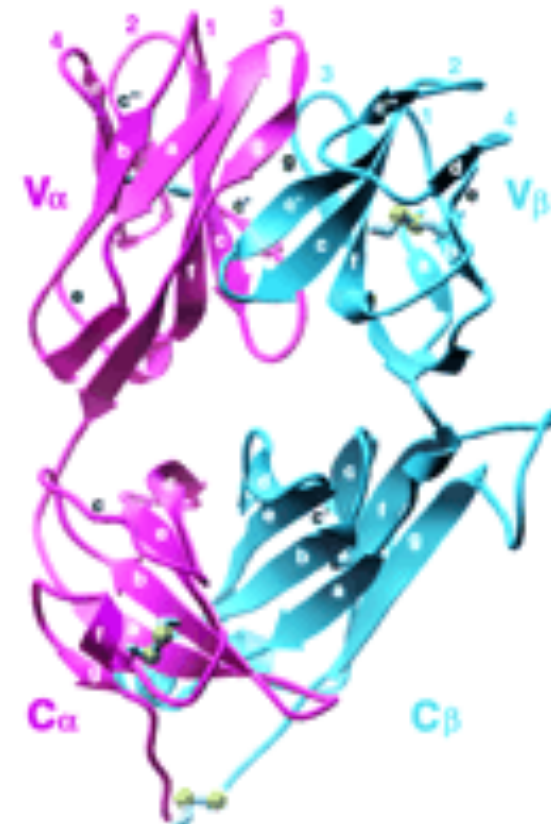
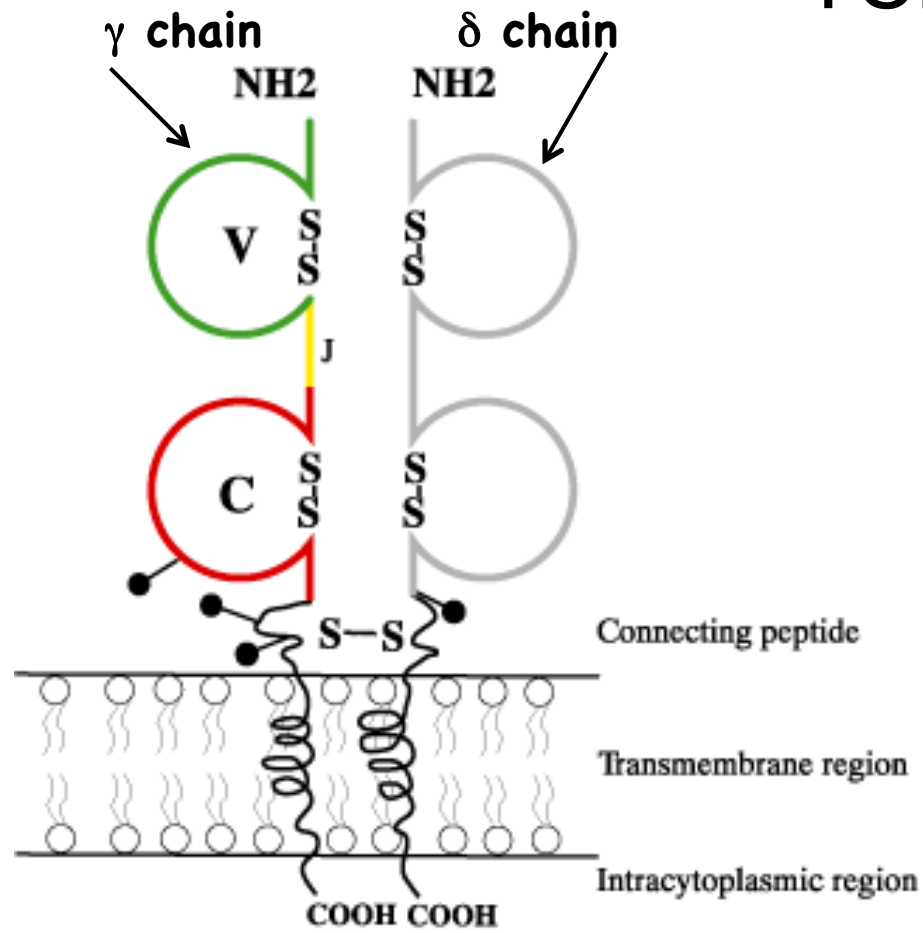
Paolo Marcatili



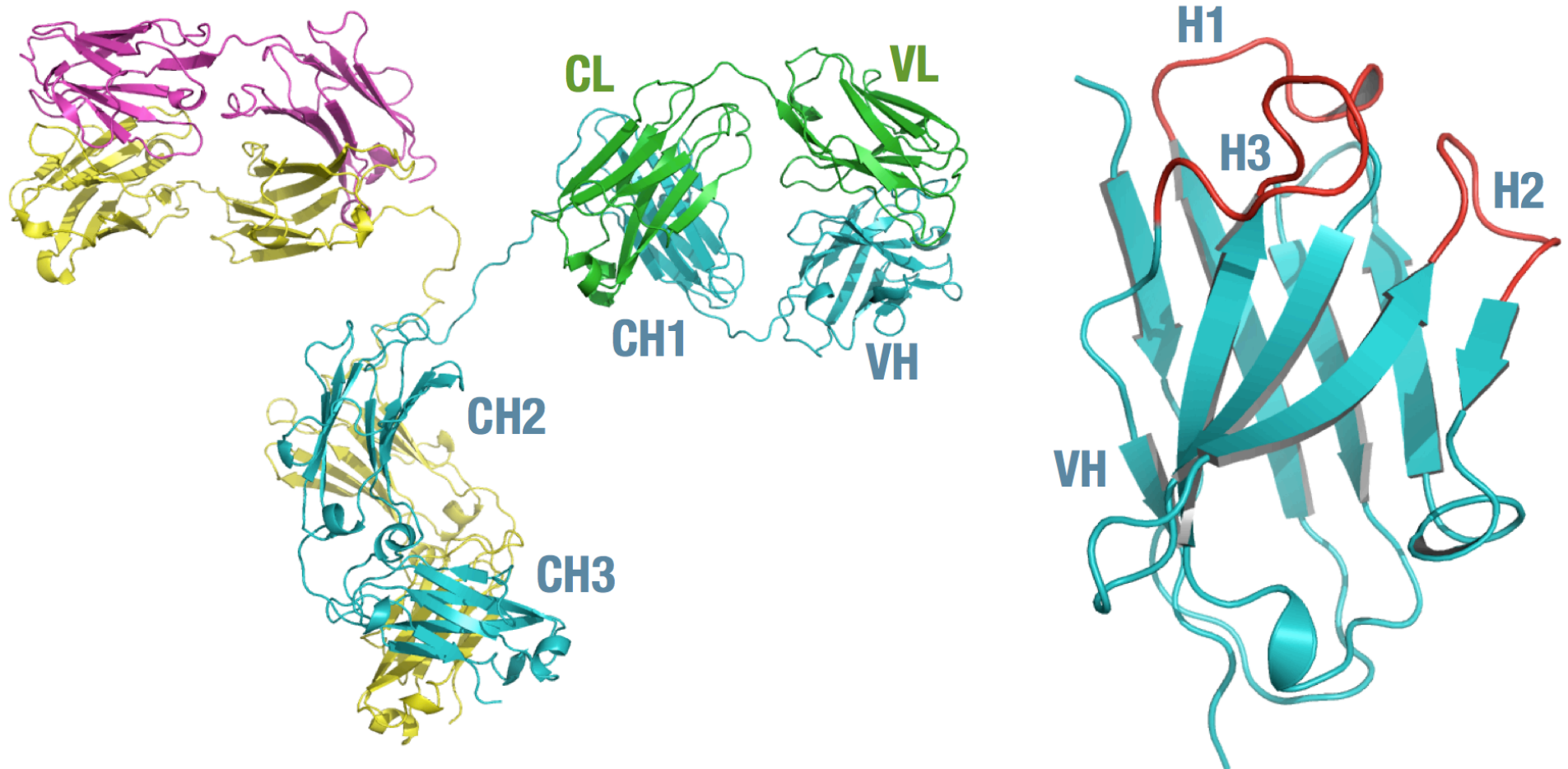
# The Ig superfamily



# TCRs

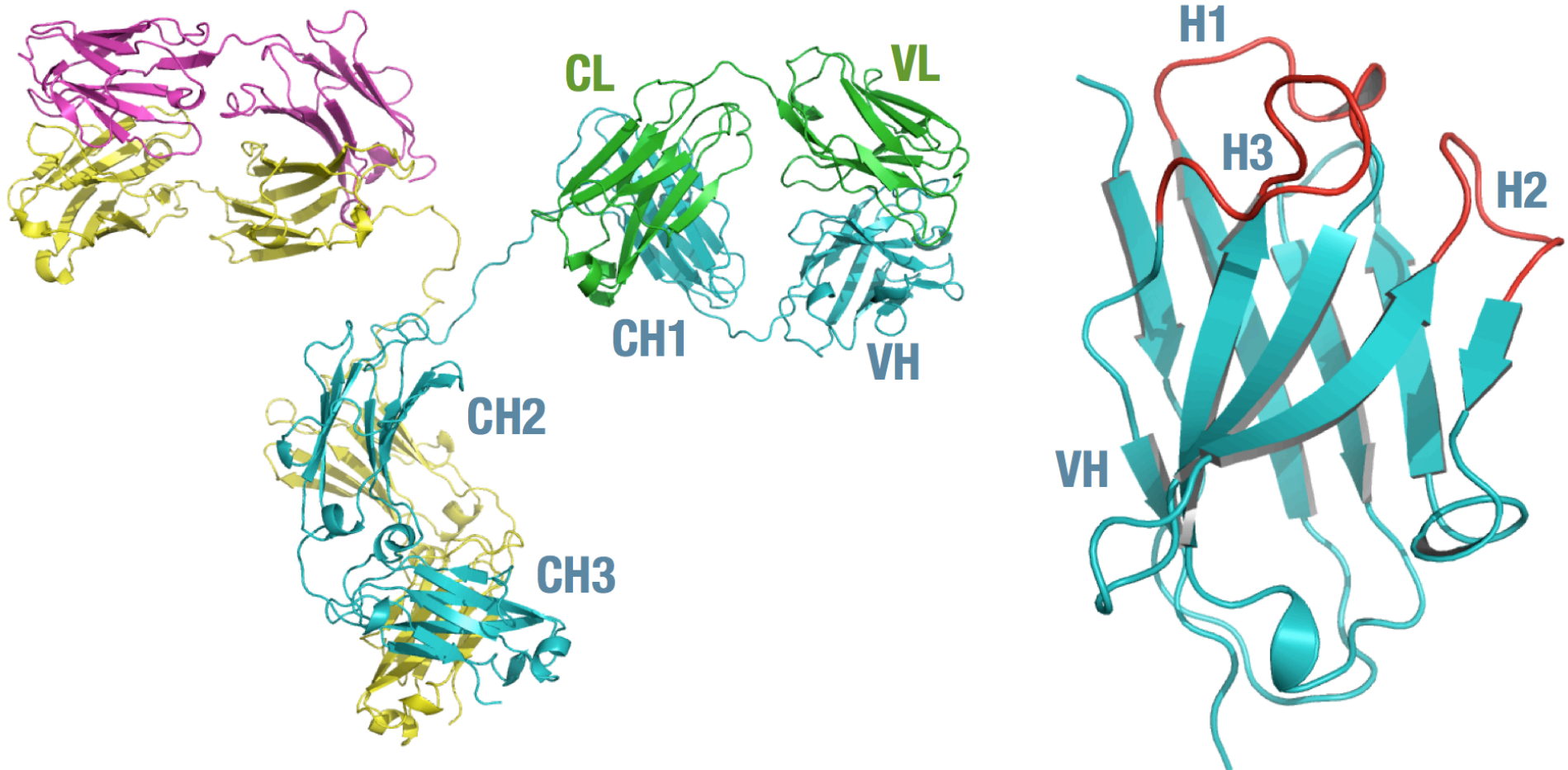


# Basic properties

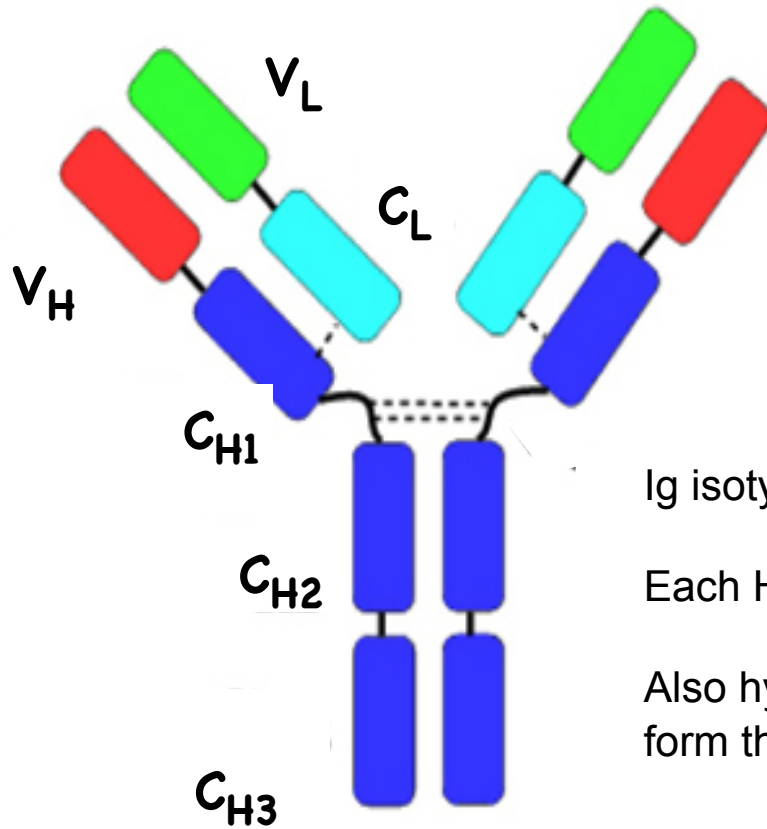


# Basic properties

- Beta sandwich
- 2 light/alpha, 2 heavy/beta
- held together by disulphide
- 1 variable domain, 1 to 5 constant domains
- variable = antigen specificity
- constant = effector function



## Ab STRUCTURE, cont.



5 types of heavy chain:  $\mu$ ,  $\delta$ ,  $\gamma$  (4 subtypes),  $\epsilon$ ,  $\alpha$ .  
H chain defines class (isotype) of Ig: IgM, IgD, IgG, IgE, IgA

Ig isotypes play different roles in immune responses.

Each H or L chain composed of a C and a V region.

Also hypervariable regions interspersed throughout V. These form the 3-D structure of the Ag binding site.

# CLASSES AND SUBTYPES OF HUMAN Ig

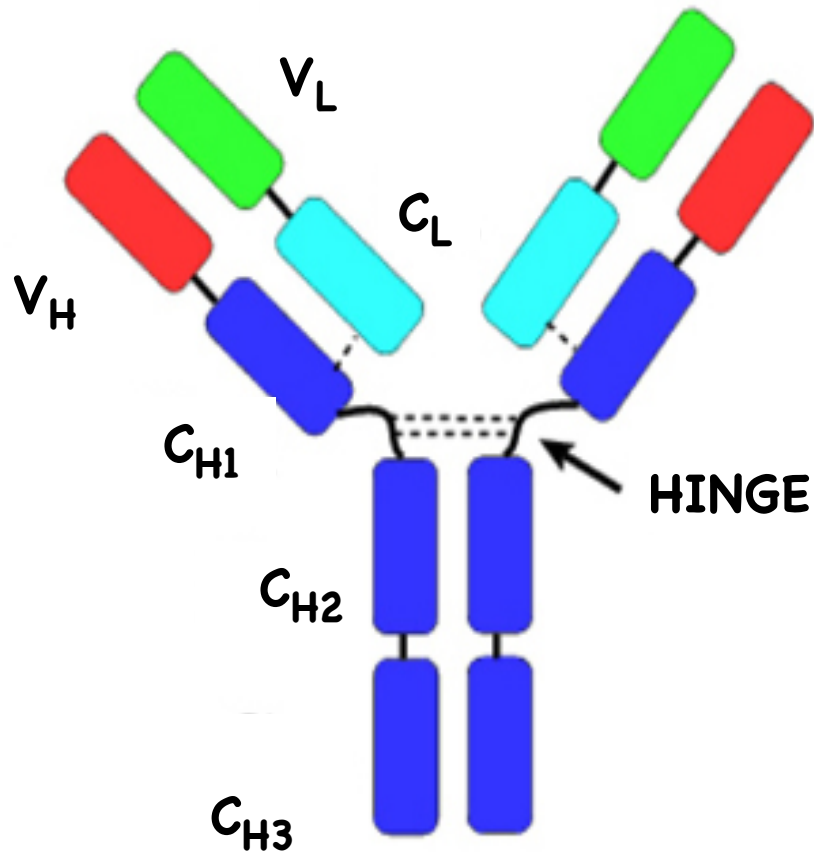
## Ig class or subtype

property	IgM	IgG1	IgG2	IgG3	IgG4	IgA1,2	IgE	IgD
form	pentamer	mono	mono	mono	mono	dimer	mono	mono
serum level (mg/ml)	1.5	9	3	1	0.5	3.5	0.00005 <sup>#</sup>	0.03
Complement activation	+++	+++	+	+++	-	-	-	-
Placental transfer	-	+	+	+	+	-	-	-
macrophage (Fc receptor) binding	-	+	-	+	-	-	-	-
present in external secretions	mucus* etc	milk *	milk *	milk *	milk *	mucus* etc	-	-
mast cell/ basophil degranulation	-	+*	-	-	-	-	+++*	-

\*(species-dependent)

# value is for non-allergic individuals

## Ab STRUCTURE, cont.



H have 4 domains

Amino terminal variable domains ( $V_H$ ) at the tips of the Y

3 constant domains:  $CH1$ ,  $CH2$ , and the COOH terminal  $CH3$ , at the base of the stem

Switch connects the constant and variable regions

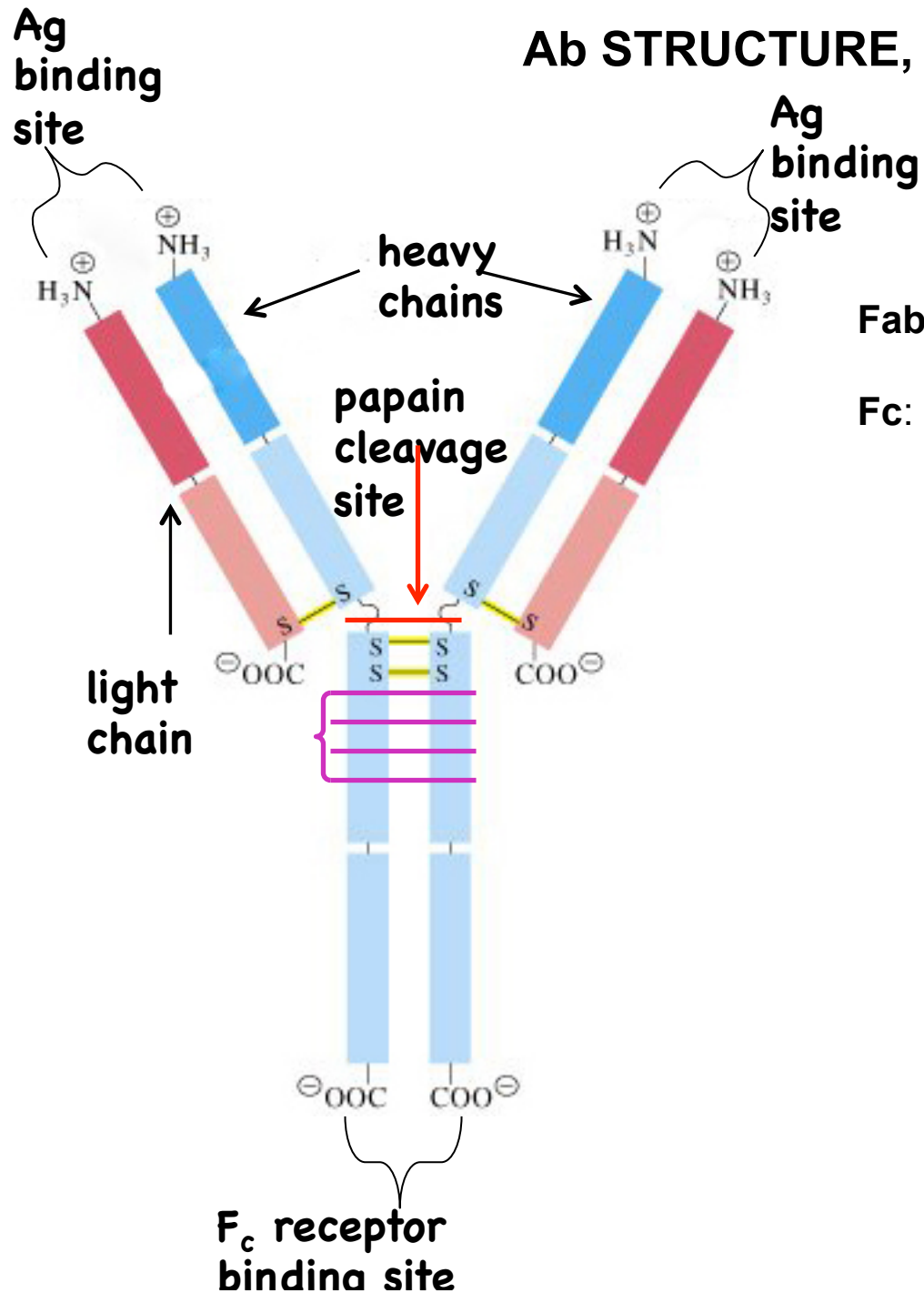
Hinge connects  $CH2$  and  $CH3$  (the  $F_c$  fragment) to the rest of Ig  $F_{ab}$  fragments).

L chains - 2 domains, variable ( $V_L$ ) & ( $C_L$ ), connected by a switch.

Individual B cells synthesize Abs of a single specificity (same H, same L).



## Ab STRUCTURE, cont.

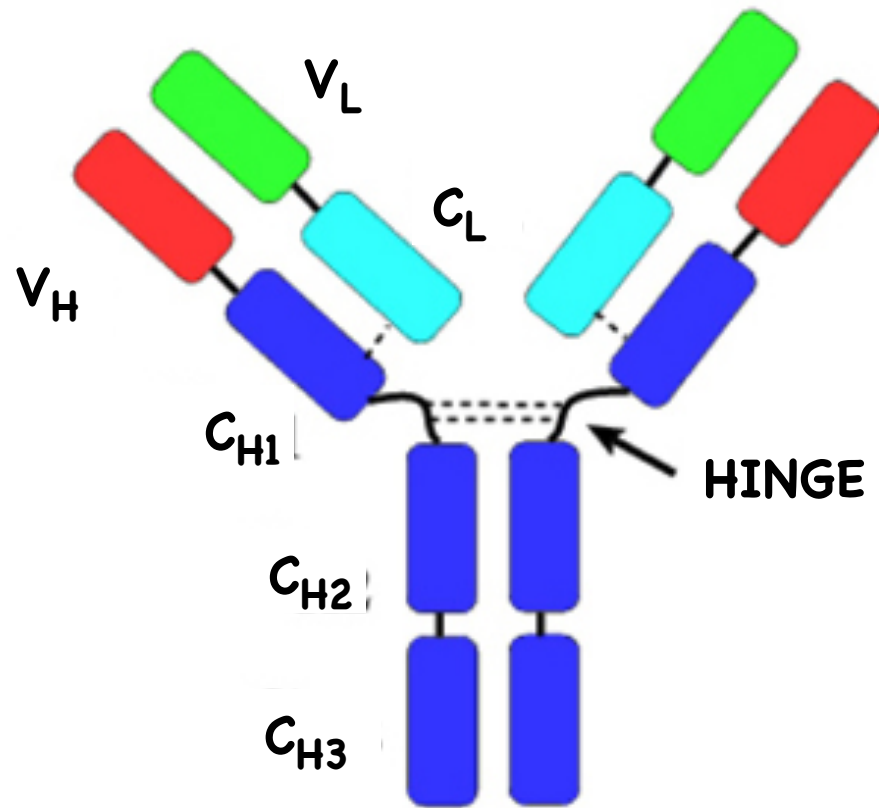


Cleavage of Ig with papain yields 3 fragments of 2 types:

**Fab:** has antigen (Ag) binding site.

**Fc:** C = crystal; this portion of Ig 1st to be crystallized. F<sub>c</sub> binds to cell surface F<sub>c</sub> receptors, augmenting immune responses (opsonization, e.g.).

## Ab STRUCTURE, cont.



V of both L and H Ig chains contain 3 hypervariable regions, or complementarity-determining regions (CDRs).

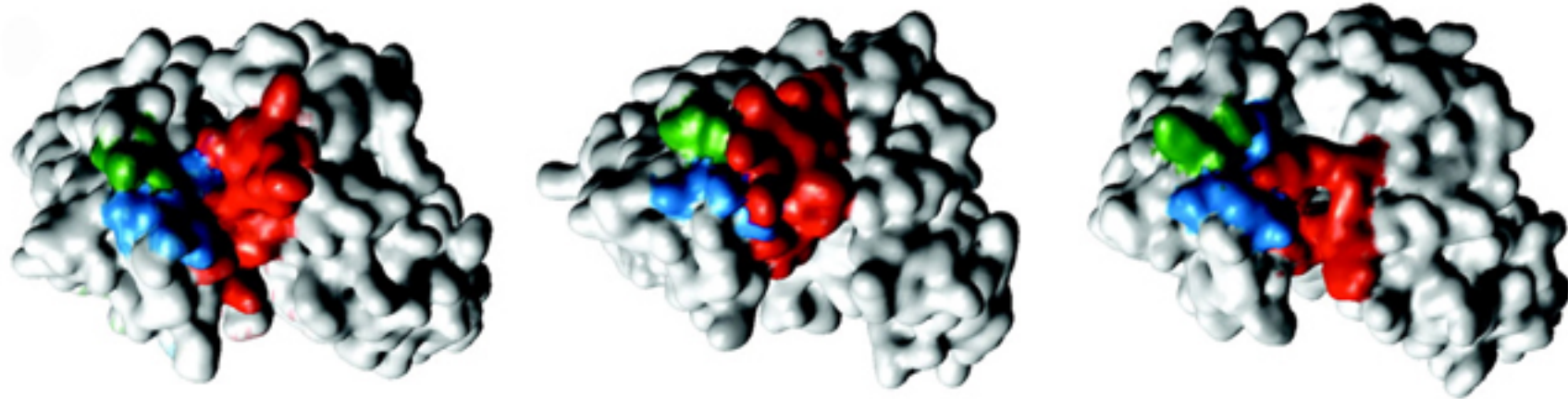
CDRs = loops connecting  $\beta$  strands in Ig fold

Residues in CDRs vary from one Ab to the next, imparting Ag specificity to each Ab.

V<sub>L</sub> and V<sub>H</sub> domains at tips of Abs packed so that 6 CDRs (3 on each) form surface for Ag-specific binding.

Residues from all 6 CDR's (V<sub>L</sub> CDR1, CDR2, CDR3 & V<sub>H</sub> CDR1, CDR2, CDR3) project from distal surface of Ab tip to bind Ag.

## EFFECT OF SEQUENCE ON Ab BINDING SITE & Ag SPECIFICITY



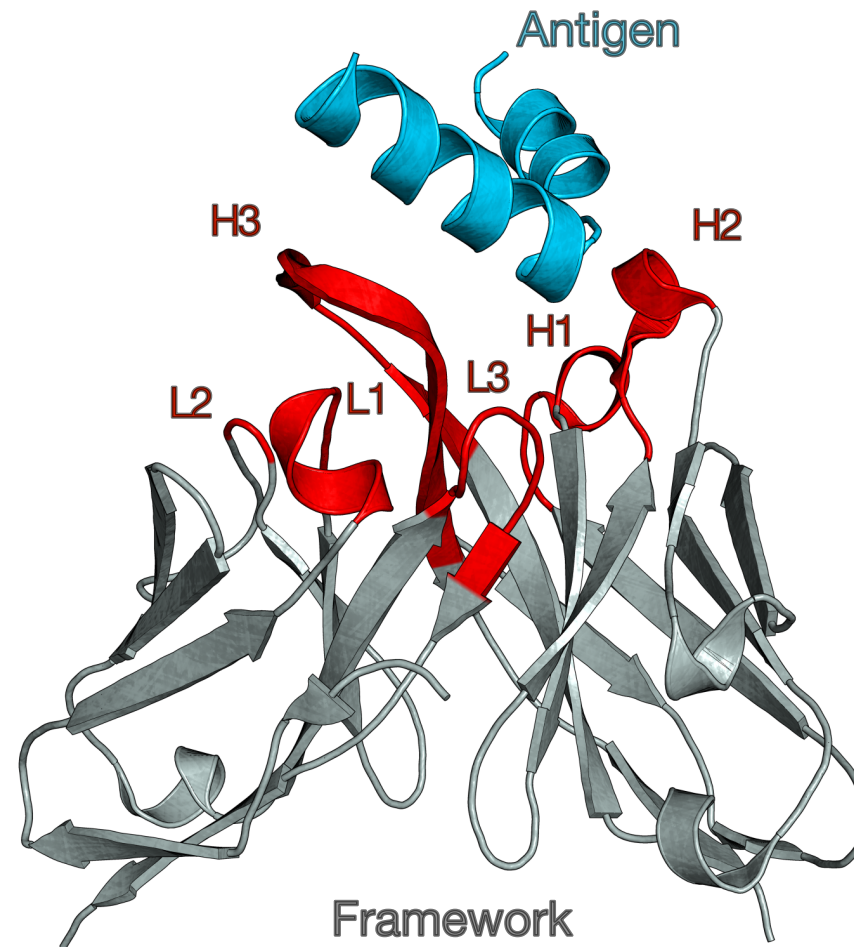
from Lantto, J., et al. J. Biol. Chem. 2002; 277: 45108-45114.

red = CDHR3; blue = CDHR2; green = CDHR1

CDHR = complementarity-determining region of H chain  
(hypervariable region)

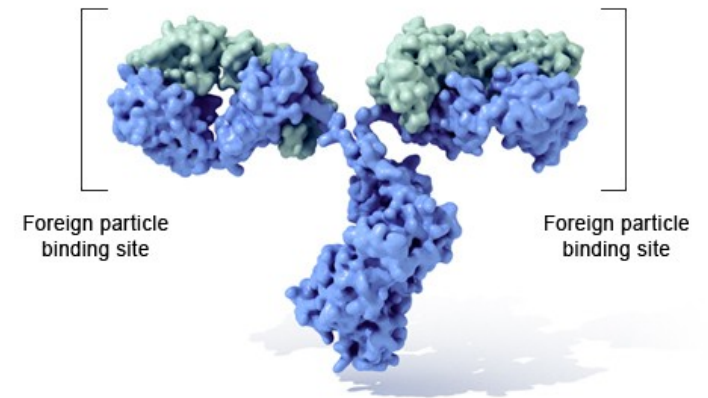
Effect of insertions/deletions on Ag binding

# Antibody Binding



# Immunoglobulins

Immunoglobulin G (IgG)

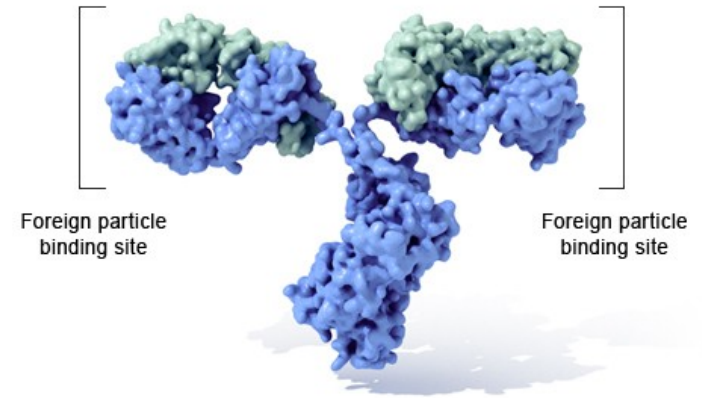
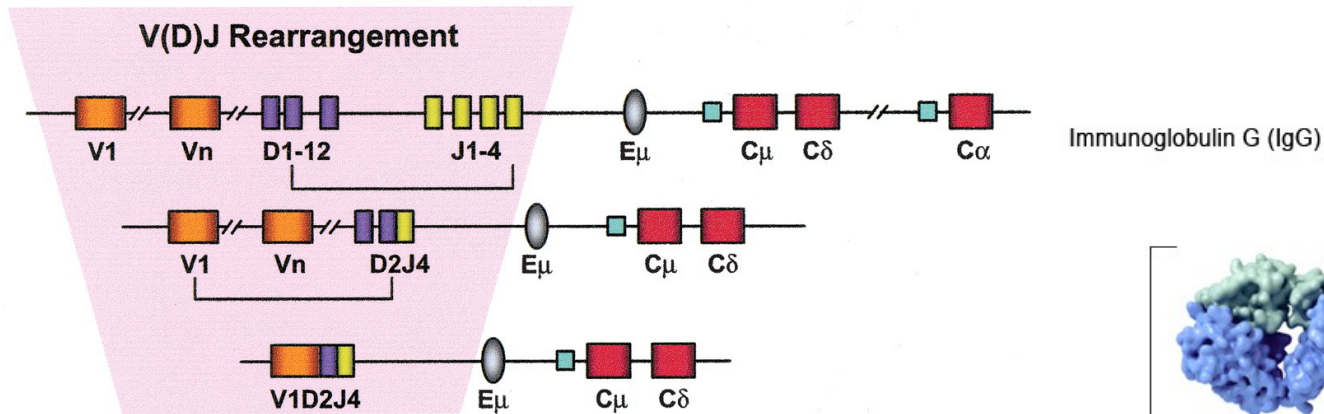


U.S. National Library of Medicine

The generation of antibody diversity  
through somatic hypermutation and  
class switch recombination  
Ziqiang et al., G&D 2004

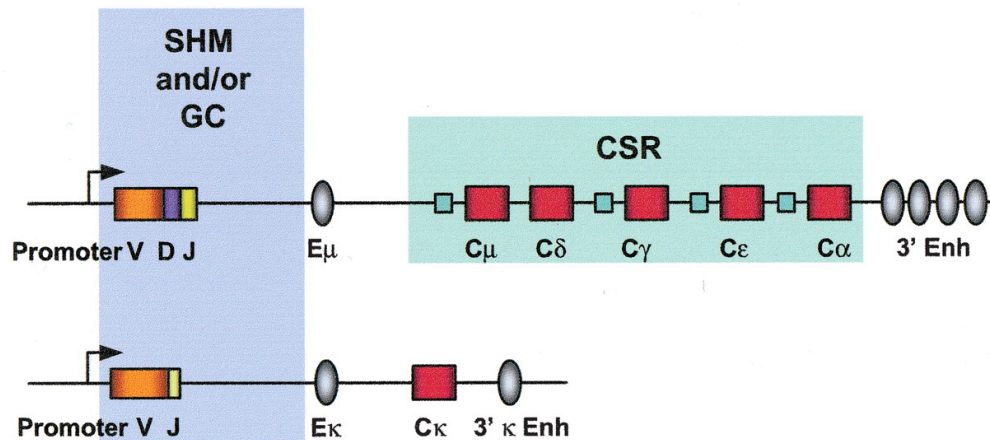
# Immunoglobulins

A



U.S. National Library of Medicine

B



The generation of antibody diversity through somatic hypermutation and class switch recombination  
Ziqiang et al., G&D 2004

# Hypervariable regions

```
EVQLVESGGGLVQPGGSLRLSCAASGFTFSN----YEMNWVRQAPGKGLEWISYISN----GDNTIYYADSVKGRFTISRDSAKNSLYLHMHLRAEDTAVYYCARGDYGGNGYFYYAM-----DVWGQGTITVTVSS
EVQLVESGGGLVQPGSLRLSCTASGFTFGD----YAMSWFRQAPGKLEWVGFIRSKA--YGGTTEYAASVKGRFTISRDDSKSIAYLQMNSLKTEDTAVYYCTRDRIGNYDFWSGYYTG-----GYWGQGTITVTVSS
-VQLVQSGAEVKKPGSSVKVSKASGGTFSS----YAISWVRQAPGQGLEWMGGIIP---IFGTANYAQKFQGRVTITADESTAYMELSSLRSEDVAVYYCARVWGGSGSYIWF-----DPWGQGTITVTVSS
QVQLVQSGAEVKKPGASVKVSKASGYTFTG----YYMHWVRQAPGQGLEWMGWINP----NSGGTNYAQKFQGRVTMTSDTSISTAYMELSLRSDDTAVYYCAREQWLVEHYF-----DYWGQGTITVTVSS
-VQLVQSGAEVKKPGSSVKVSKASGGTFSS----YAISWVRQAPGQGLEWMGGIIP---IFGTANYAQKFQGRVTITADESTAYMELSSLRSEDVAVYYCATKNDFWSGYYEGYYYYYIM-----DVWGKGTITVTVSS
QVQLVQSGAEVKKPGSSVKVSKASGGTFSS----YTISWVRQAPGQGLEWMGRIIP---ILGIANYAQKFQGRVTITADKSTAYMELSSLRSEDVAVYYCATNYDFWSGY-----PYWGQGTITVTVSS
EVHVESGGGLVQPGGSLRLSCAASEFTFDR----YWMWVRQAPGKLEWVNIKE----DGSEKKYVDSVRGRFTISRDNKNSLYLQMNSLRAEDTAVYYCARGAYF-----GYWGQGTITVTVSS
```



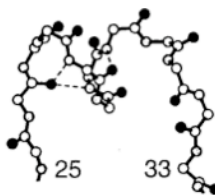
HV loops (ABS) model unreliable?



# Canonical Structures

Repertoire of discrete conformations

Key residues



Torsion angles:

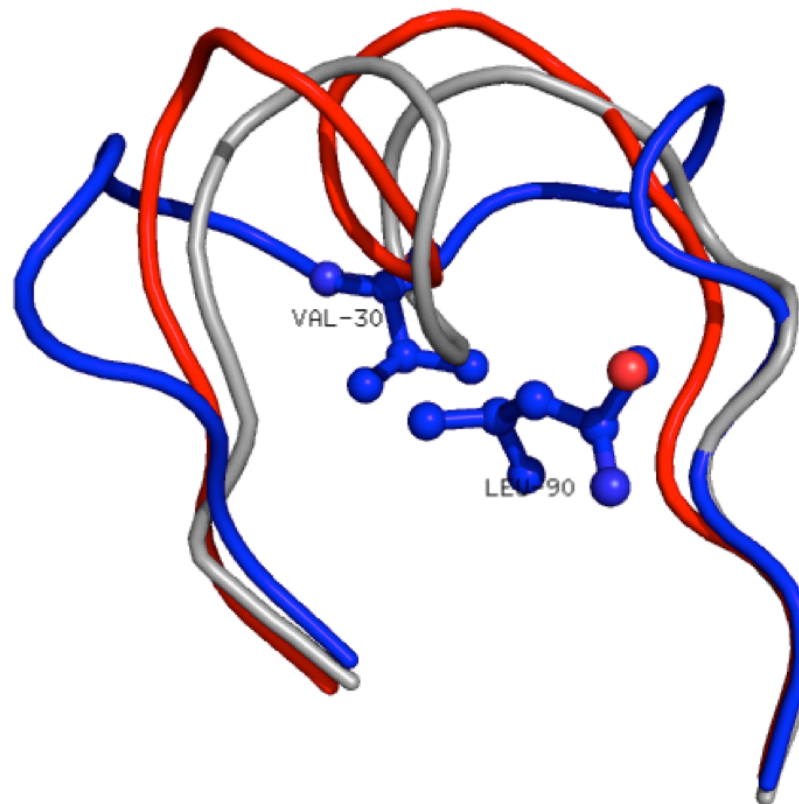
Residue	Angles		
25	$\phi$	-120	$\psi$ -135
26	$\phi$	-149	$\psi$ -180
27	$\phi$	-55	$\psi$ -23
28	$\phi$	-84	$\psi$ -19
29	$\phi$	-127	$\psi$ -88
30	$\phi$	-61	$\psi$ -38
30a	$\phi$	-74	$\psi$ 7
30b	$\phi$	-116	$\psi$ 26
30c	$\phi$	70	$\psi$ 37
31	$\phi$	-109	$\psi$ 148
32	$\phi$	-71	$\psi$ 157
33	$\phi$	-119	$\psi$ 125

## Hydrogen Bonds:

26	O	- - -	N	29
26	O	- - -	N	30a
29	O	- - -	N	30b

Structure	PDB file	Residues
NEW	7fab	GSSSNIGAGHNV

Figure 7.  $\lambda$  L1 canonical structure 2.





Loop	CS	Length	Constraints
κL1	1	6	29 VIL
	2	7	29 VIL
	3	13	29 VIL
	4	12	29 VIL
	5	11	29 VIL
	6	8	29 VIL
κL2	1	3	
κL3	1	6	90 QNH 95 P
	2	6	90 Q 94 P
	3A	5	90 Q 91/92 G 96 P
	3B	5	90 Q 91/92 ST 96 P
	4	4	90 Q
	5	7	90 Q 95A P
	6	5	90 Q 94 L
	7	5	94 P
λL1	8	8	90 Q 95A P
	1	10	25 G
	2	11	25 G 31 FHY 66 K 90 S
	3	11	66 L 90 L
	4	8	28 VIL 66 ST
	5	10	25 R 28 G
	6	11	25 G 31 ND 66 K 90 S
	7	8	28 VIL 66 N
λL2	8	9	
	1	3	
λL3	2	7	
	1 form A	6	
	1 form B	6	
	1 form C	6	
	1 form D	6	
	2	8	
	3	7	92 D 95 ST
	4	7	95 not ST

H1	1	7	
	2	8	
	3	9	
H2	1	3	
	2	4	71 AVL
	3	4	71 RK
	4	6	
H3	Bulged	>10	94 RK 101D
	Non-bulged	>10	94 not RK
	Short	10	

# Canonical Structure prediction

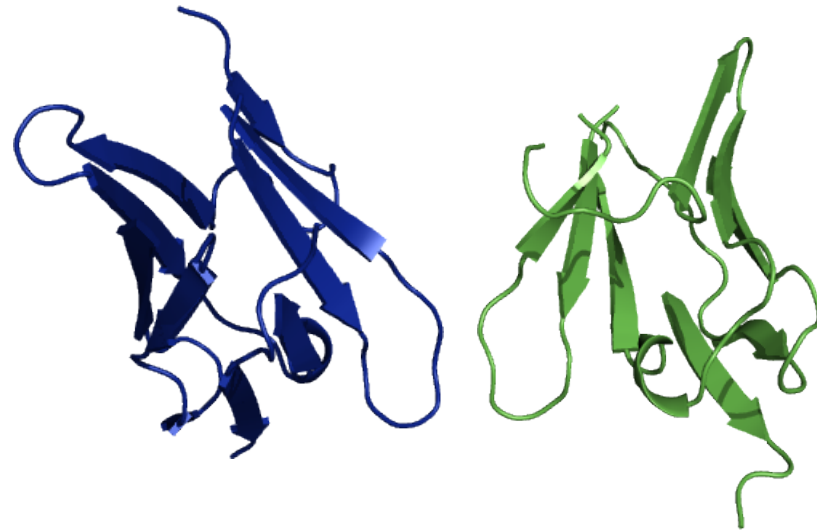
Given an antibody sequence

we can predict the CS of each loop (H3 is kind of an exception)

[illegible]

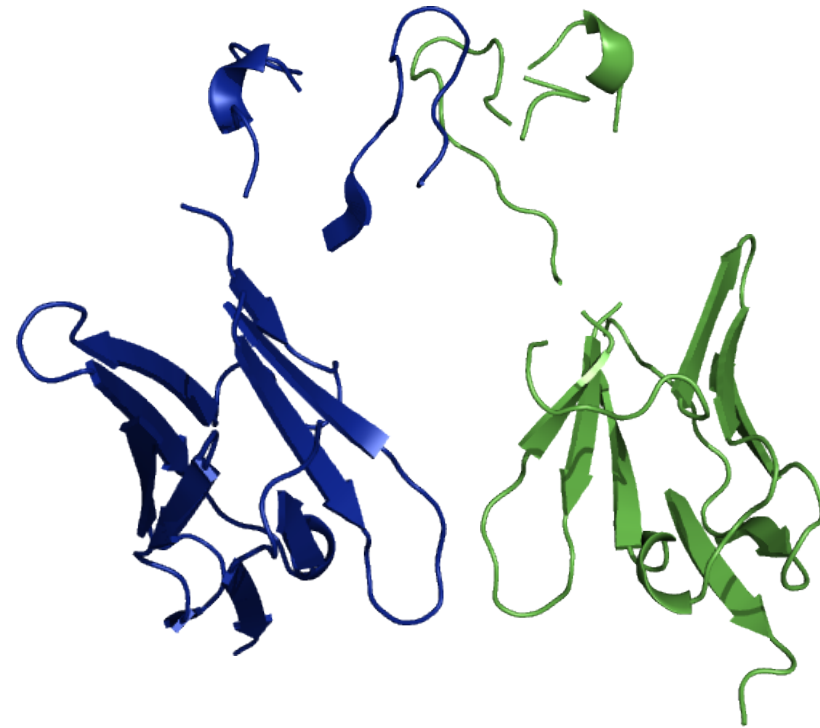
# PIGS

**Frameworks** (L & H) selection based on %id



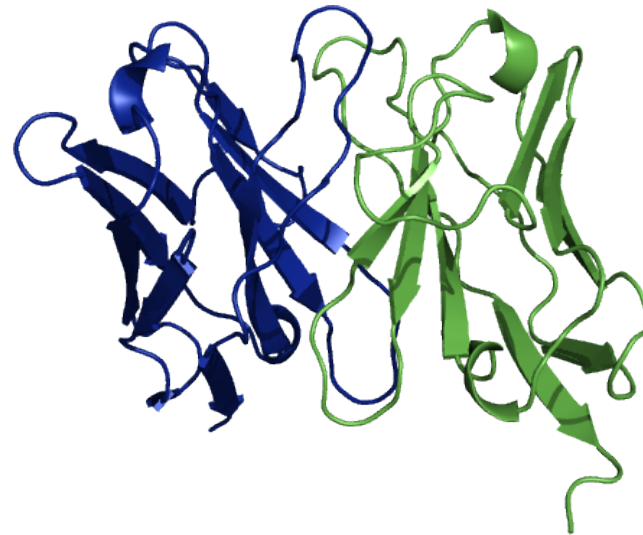
# PIGS

Loop selection (L1-3, H1-3) based on CS + %id



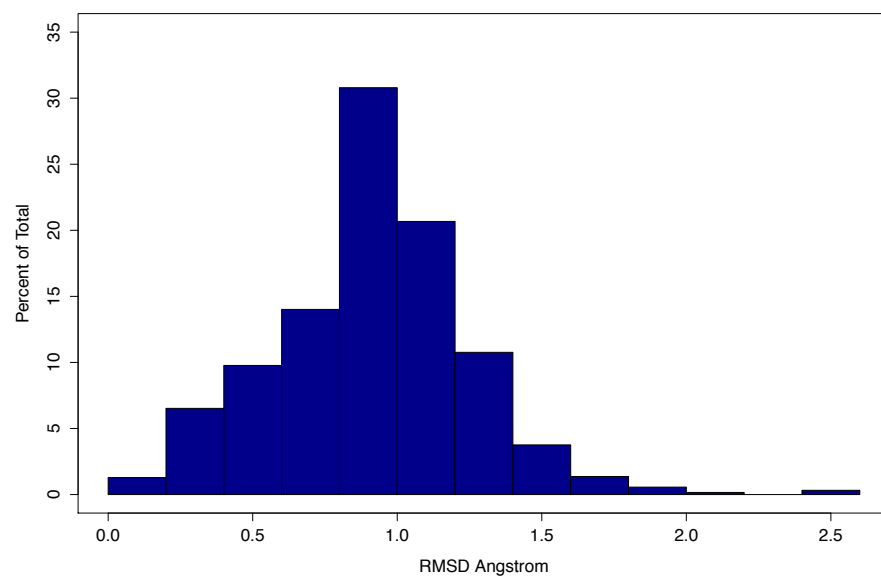
# PIGS

VL-VH **packing**  
**Superpose loop** stems  
**Side Chains** (SCWRL)

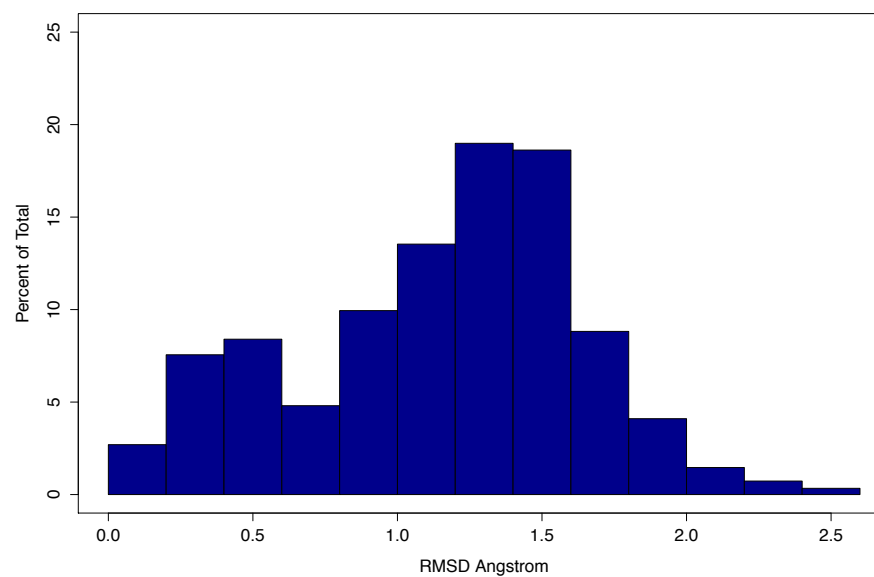


# Assessment

Whole Fv



ABS (all CDRs)



# H3 Prediction

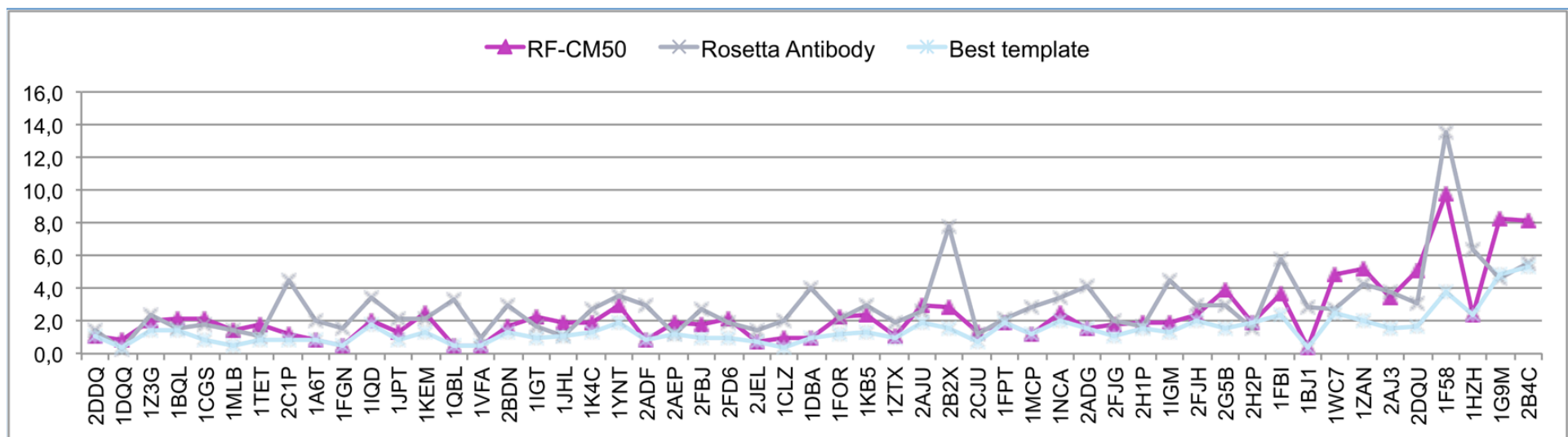
H3 HCDR3) is:  
The longest loop (up to 25 residues)  
(Almost) no CS  
The most flexible  
The most important for recognition  
The hardest to predict



# H3 accuracy

Good ( $\sim 1\text{\AA}$ ) for short loops

Bad (up to  $6\text{\AA}$  and more) for long loops





# TCR models

We have > 1500 solved antibodies

TCR: ~50

No method currently available!

# Problems in modeling

H3 models are unreliable

Bias on Human and Mouse antibodies

Role of glycosilation

Signal transduction

# Take home

The antibody fold is one of the most common  
beta-sandwich

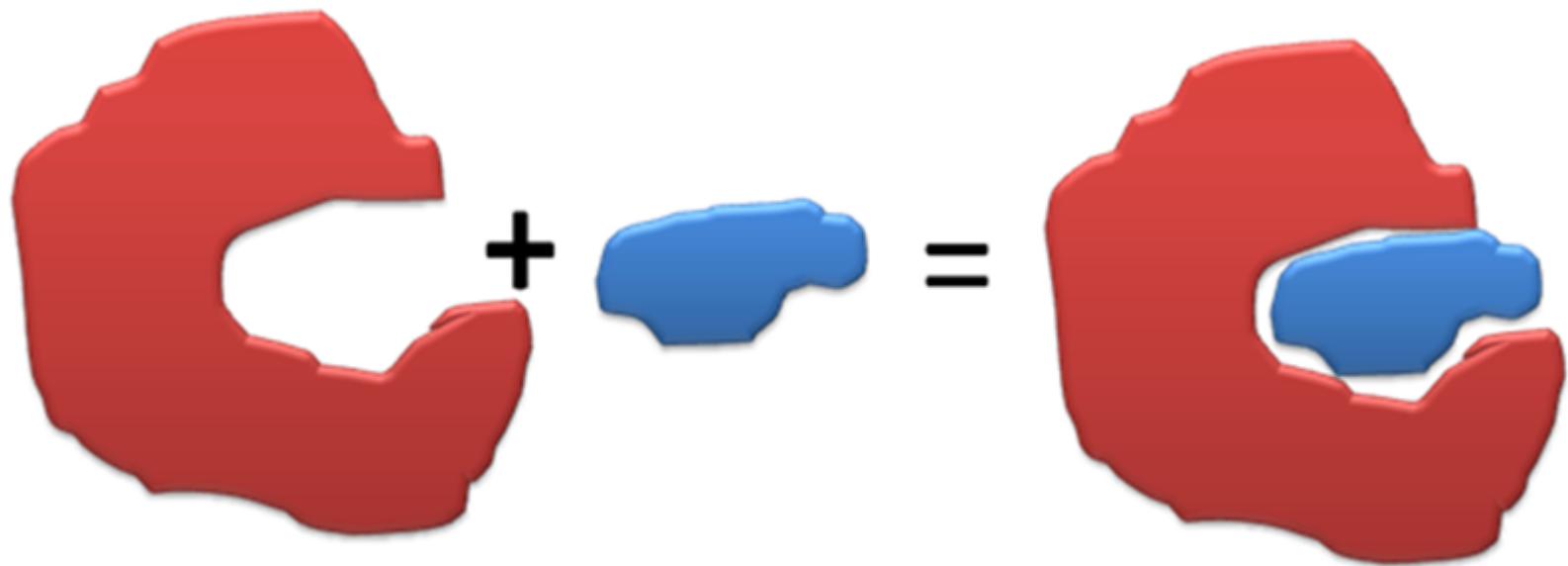
CDRs are located on loop regions  
extremely variable in sequence, not in structure (backbone!!!)

Canonical structure method

Models are extremely reliable (with the exception of H3)

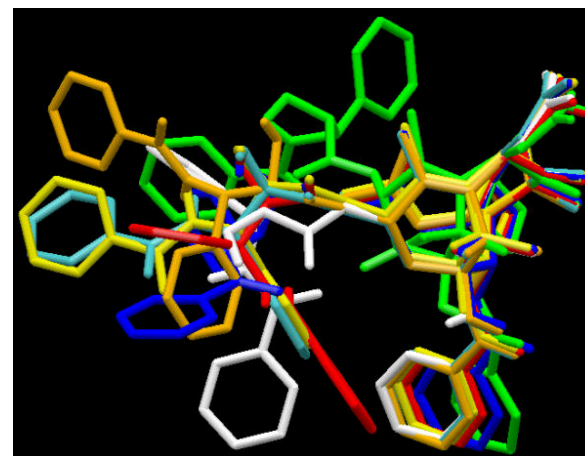
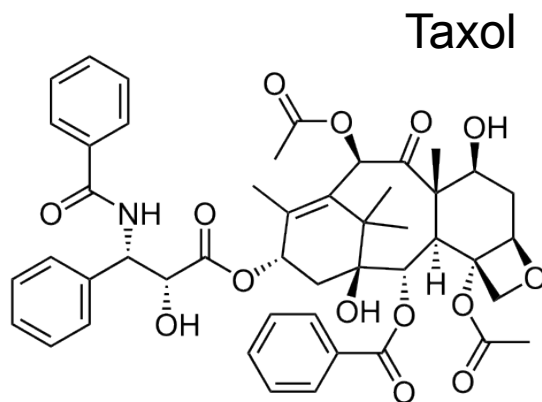
# Rigid Body Docking

---



# Degrees of freedom

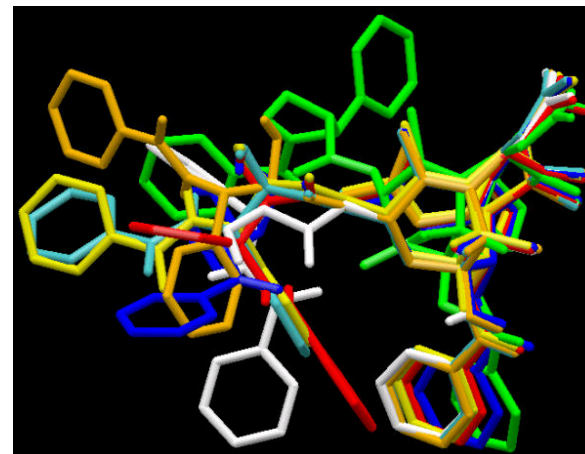
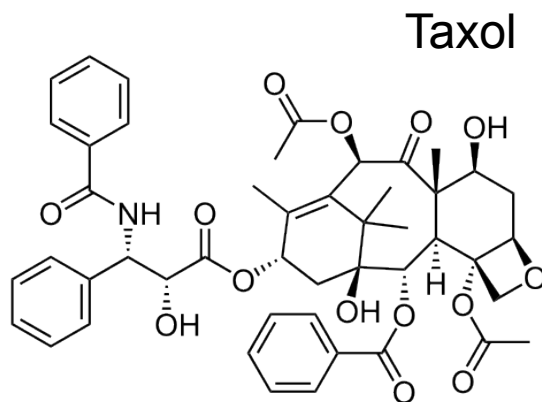
- Conformations are different three-dimensional structures of molecules that result from rotation about single bonds
- If the torsion angles are incremented in steps of  $30^\circ$ , this means that a molecule with 5 rotatable bonds will have  $12^5 \approx 250K$  conformations
- Having too many rotatable bonds results in “combinatorial explosion”
- Also ring conformations



# Degrees of freedom

Find few stable conformers, perform multiple docking

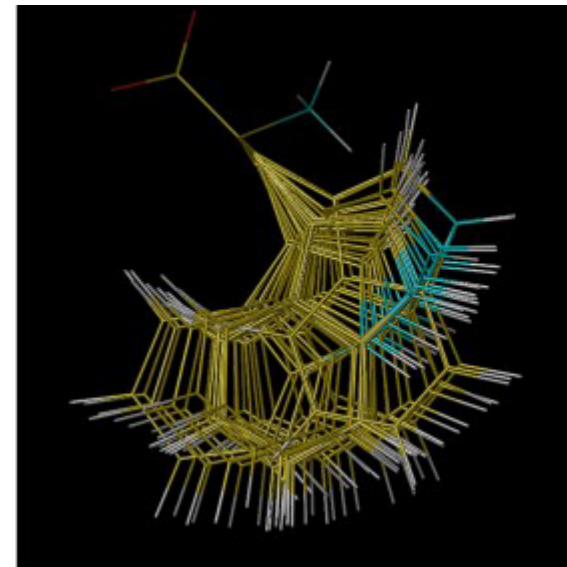
Run Rigid-Body Docking, refine the final solutions allowing flexibility



# Induced Fit

---

the protein may deform slightly to  
accommodate different ligands  
(**ligand-induced fit**)  
protein side chains in the active  
site may adopt different  
conformations

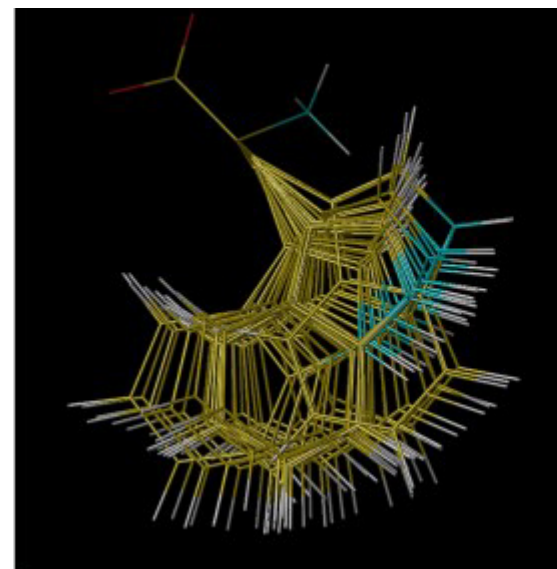


# Induced Fit

---

the protein may deform slightly to accommodate different ligands  
(**ligand-induced fit**)  
protein side chains in the active site may adopt different conformations

- Some docking programs allow protein side-chain flexibility
  - For example, selected side chains are allowed to undergo torsional rotation around acyclic bonds
  - Increases the search space





# Energy functions

---

## Forcefield-based

Based on terms from molecular mechanics  
forcefields

GoldScore, DOCK, AutoDock

## Empirical

Parameterised against experimental binding  
affinities

ChemScore, PLP, Glide SP/XP

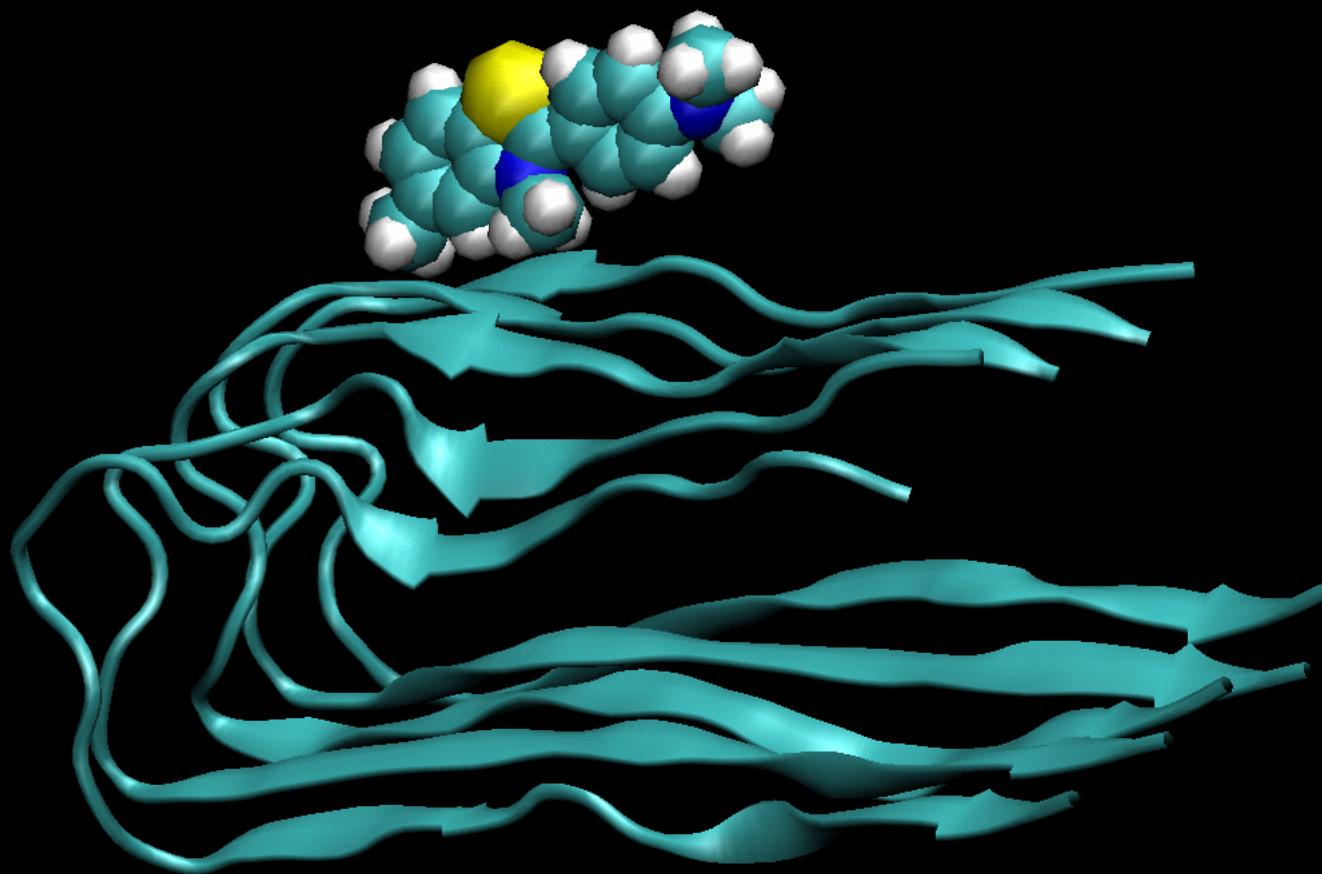
## Knowledge-based potentials

Based on statistical analysis of observed  
pairwise distributions

PMF, DrugScore, ASP

# Free Docking

---

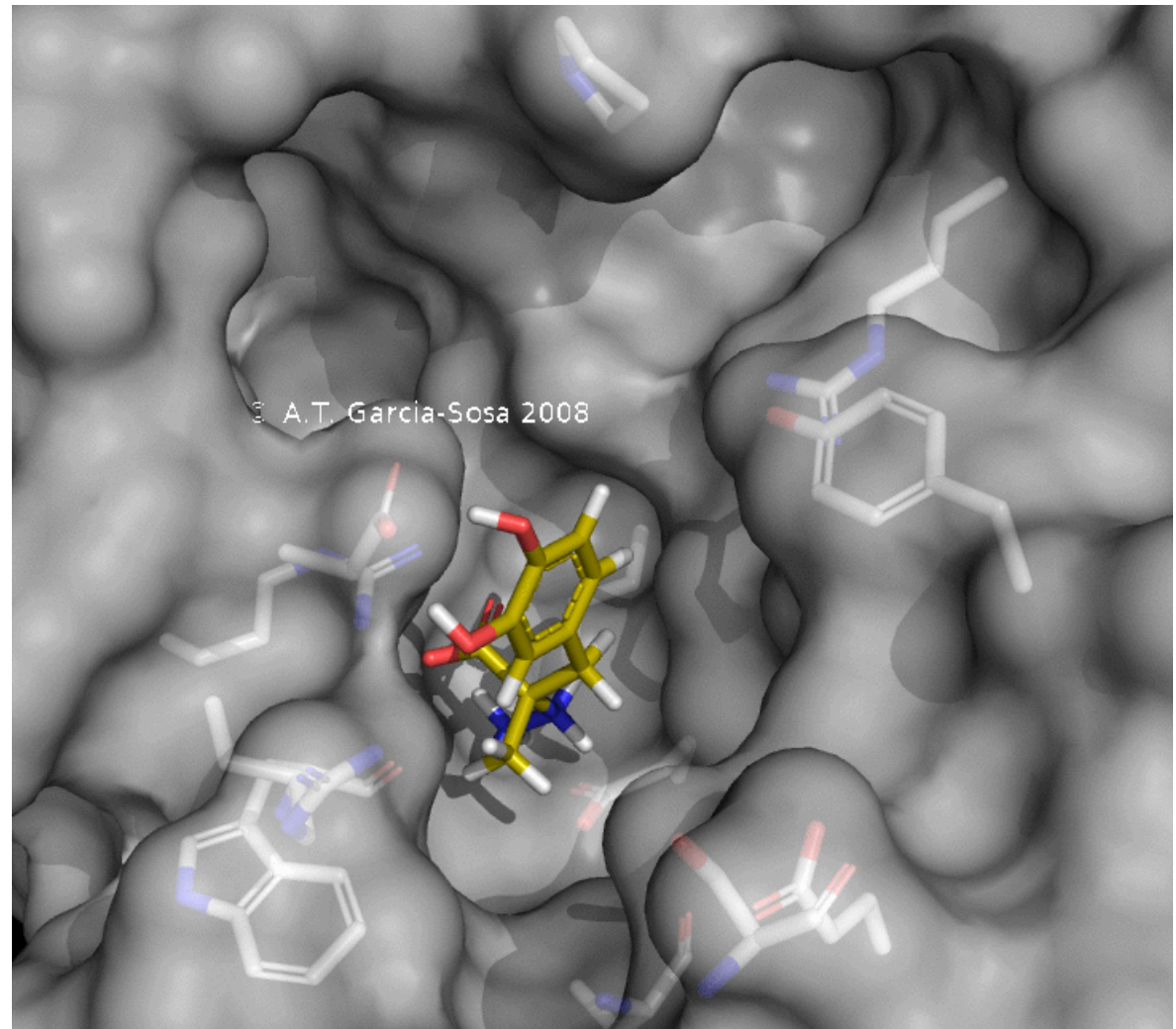


# Constrained Docking

Known Binding site

Compare different ligands

Biased results



# Docking Pipeline

---

Prepare protein (minimization, Box)

Run >10.000 simulations

Select 1.000 with best energy

Rescore

Cluster

# Docking Pipeline

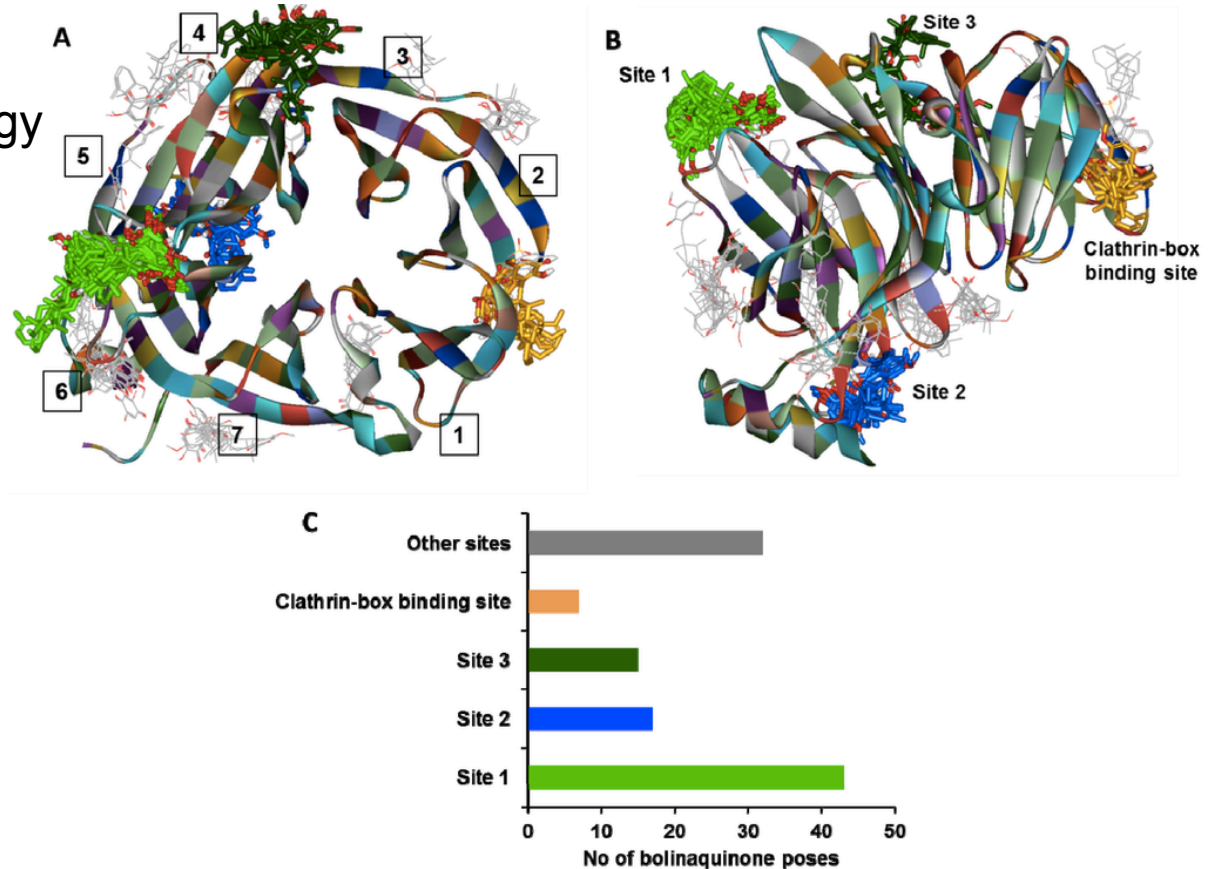
Prepare protein (minimization, Box)

Run >10.000 simulations

Select 1.000 with best energy

Rescore

Cluster



# Take Home 2

---

Docking on solved structures or models

Flexibility (induced Fit) vs complexity

Free vs Constrained Docking

Run many simulations & cluster

No Scoring Function is perfect!